DOCUMENT RESUME

ED 389 729 TM 024 285

AUTHOR	Johnson, Colleen Cook; Rakow, Ernest A.
TITLE	Effects of Violations of Data Set Assumptions When Using the Analysis of Variance and Covariance with Unequal Group Sizes.
PUB DATE	11 Nov 94
NOTE	27p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994). Tables contain small print.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Analysis of Covariance; *Analysis of Variance; *Educational Research; Monte Carlo Methods; Research Methodology; *Robustness (Statistics); *Simulation; *Statistical Significance
IDENTIFIERS	*Group Size; Heterogeneity of Variance; Kurtosis; Skew Curves; Type I Errors; *Violation of Assumptions

ABSTRACT

This research explored the degree to which group sizes can differ before the robustness of analysis of variance (ANOVA) and analysis of covariance (ANCOVA) are jeopardized. Monte Carlo methodology was used, allowing for the experimental investigation of potential threats to robustness under conditions common to researchers in education. The effects of unequal group sizes were explored under the following data set conditions: (1) heterogeneity of group variances; (2) skew; (3) kurtosis; and (4) in ANCOVA, heterogeneity of regression slopes. Two independent sets of simulations were conducted, one using a total group of 90 and the other a total of 60. Experimentation was limited to simulations using three groups, with the total number divided in a systematic fashion. Results of these studies produced results consistent with previous research. In the analyses having homogeneity of group variances, the only simulations that emerged as statistically significant from the theoretical F test were those that had a large degree of difference in group numbers and unequal regression slopes. No significant differences emerged in simulations with near equal numbers. Tables in this paper document the simulation results, but also offer the research practitioner some idea of the true risk of Type I error in such situations. (Contains 7 tables and 24 references.) (SLD)

EFFECTS OF VIOLATIONS OF DATA SET ASSUMPTIONS WHEN USING THE ANALYSIS OF VARIANCE AND COVARIANCE WITH UNEQUAL GROUP SIZES

Colleen Cook Johnson, Ed. D.

Department of Sociology and Social Work, University of Memphis

Ernest A. Rakow, Ph.D.

Department of Counseling, Educational Psychology and Research, University of Memphis

INTRODUCTION

The analysis of variance (ANOVA) statistical procedure has been found to be the most widely used technique in several major educational journals, and along with its logical extension - the analysis of covariance (ANCOVA) - is widely used in psychology and the social sciences as well (Elmore and Woehlke, 1988; Goodwin and Goodwin, 1985; Halpin and Halpin, 1988). Like other statistical procedures derived from the general linear model, ANOVA and ANCOVA are designed to produce results that are both statistically robust and powerful when specific mathematical conditions (i.e., statistical assumptions) are present. However, practitioners seldom find the ideal research situation where all of these conditions are met.

Previous research has, for example, found that the robustness of both the ANOVA and ANCOVA F is questionable in face of unequal group sizes and/or heterogeneity of group variances (e.g., Box, 1954; Glass et al., 1972; Johnson and Rakow, 1993; Scheffe´, 1959) - problems that researchers in these disciplines are often unable to avoid. Thus, the critical question that arises is how much difference exists between the conditions these statistical procedures were designed to handle and the actual conditions that occur in a particular researcher's situation. If the difference is within a "tolerable range," then use of ANOVA or ANCOVA should provide information that is statistically robust. However, when the data collected exceed that "tolerable range," alternatives to these parametric procedures must be considered.

Theoretical mathematics can be used to define the general nature of the problems that arise when ANOVA or ANCOVA are used inappropriately, however it is unable to provide us with the precise limits of this "tolerable range." Therefore, this research uses Monte Carlo methodology to provide supplemental information which cannot be obtained using theoretical mathematics. Using Monte Carlo methodology, it is possible to numerically define the degree of tolerance (i.e., robustness) that ANOVA and ANCOVA have in practice.

The Goal of this Research

This research is designed to examine the effects of unequal group sizes on the robustness of the ANOVA and ANCOVA statistical procedures. Specific questions this research is designed to address include the following: (1) How different can the group sizes be without jeopardizing the results? (2) Does the "tolerable range" in group sizes change when violations of data set assumptions are also present? (3) How do unequal n's affect the true level of significance (that is, the true risk of making type I error) in a given research situation?

REVIEW OF THE LITERATURE

The t-test for two independent variables tests for mean differences between two groups. The oneway, fixed-effects ANOVA is a logical extension of this t-test, testing for mean differences

between two or more groups. In oneway ANOVA, F tests the ratio of explained variance in the dependent variable relative to the variance left unexplained. The one concomitant (i.e., one covariate) ANCOVA procedure is a logical extension of oneway ANOVA, applicable when a third, continuous variable is known to have a significant effect on the dependent while having little or no effect on the independent variable. When ANCOVA is appropriate, the researcher's goal is to probe the effects of the independent variable on the dependent variable after removing the influence of the concomitant. To do this, ANCOVA first removes all variation in the dependent variable that is a function of the concomitant. Using these "adjusted scores," ANCOVA then effectively reanalyzes the data for mean differences between the groups.

According to Cochran (1957) and Winer (1962), the relationships found in the simple oneway ANOVA and even the simpler t-test for two independent samples carries over to ANCOVA. Therefore, this literature review will contain discussion of relevant theoretical and empirical research involving all of these procedures.

Statistical Procedures and the Assumptions Inherent Within Them

Statistical procedures are developed by mathematicians to be used under a specific set of mathematical conditions (i.e., assumptions about the data set being used). These conditions are designed to balance creditability (the ability to process data in a usable form) with manageability (the technique's ability to simplify many mathematical derivations and operations). Valid results are obtainable when a researcher's data set adheres to these assumptions.

However, seldom do data sets adhere perfectly to these assumptions. According to Glass, Peckham and Sanders (1972), the question that a researcher must ask in reference to the data collected is not *whether* the assumptions are satisfied, but instead, *are the violations that do occur extreme enough to compromise the results?* Box and Anderson (1955) note that to fulfill the needs of the researcher, statistical criteria should (a) be sensitive to changes in the specific factors being tested (i.e., they should be powerful) and (b) they should be insensitive to changes in extraneous factors of a magnitude likely in practice (i.e., they should be robust).

Literature Concerning the Oneway Analysis of Variance

Homogeneity of Variance

This assumption was first identified in the classical 1908 paper "The Probable Error of the Mean" by The Student (Gossett); however the publishing of empirical results in this area would wait until Hsu (1938, as cited by Scheffe´, 1959). Many published studies suggest the F test with equal sample n's is robust when faced with the single violation of unequal group variances as long as the ratio of largest to smallest group variances does not exceed three (e.g., Glass et al., 1972). However, empirical research by Johnson and Rakow (1993) found balanced ANOVA results to be

robust even when the ratio of largest to smallest group variance was five. F is not robust, however, when both group sizes and variances are unequal. When n's are unequal and two groups used, inflated type I error rates occur when the larger group is paired with the smaller variance (Box, 1954; Glass et al. 1972; Johnson and Rakow, 1993; Shields, 1978). Furthermore, Tomarken and Serlin (1986) have argued that ANOVA may not be the best choice in the presence of heterogeneity of variances, especially when many groups are to be compared. They suggest that the effects of variance heterogeneity increases with the number of groups.

Normality of the Distribution of Errors

Research dating back seven decades has investigated violations of this assumption. Perhaps among the most notable is the work of Games and Lucas (1966), who suggest that skewed distributions are a greater threat to robustness than leptokurtic or platykurtic distributions. Furthermore, they suggest that F test results may improve some for highly leptokurtic error distributions. Johnson and Rakow (1993), however, found that distributional shape did not have a significant effect on the robustness of ANOVA with equal n's.

Extension of ANOVA Assumptions to ANCOVA

The simplest form of ANCOVA (which consists of one independent, one concomitant and one dependent variable) is an extension of oneway ANOVA. According to Cochran (1957) and Winer (1962), the assumptions previously discussed in regards to ANOVA apply to ANCOVA as well, provided the concomitant variable is normal.

The Seven Assumptions of the Analysis of Covariance

Elashoff (1969) and McLean (1979, 1989) reported the following seven assumptions associated with ANCOVA: (a) the cases are assigned at random to treatment conditions; (b) the covariate is measured error-free (that is, perfect reliability of the covariate measurement); (c) the covariate is independent of the treatment effect; (d) the covariate has a high correlation with the dependent variable; (e) the regression of the dependent variable on the covariate is equal for each group; (f) for each level of the covariate, the dependent variable is normally distributed; and (g) the variance of the dependent variable at each value of the covariate is constant across groups. These assumptions can be classified as falling into one of two categories: (a) assumptions concerned with research design and sampling (methodological assumptions) and (b) assumptions concerned with the the data set and its parent population (data set violations).

Methodological Assumptions

Two assumptions deal with research design and sampling: (a) the cases are assigned to random treatments and (b) the covariate has perfect reliability. Concerning randomization, Evans and Anastasio (1968) distinguished three separate scenerios: (a) individuals are assigned to

groups at random after which treatments are randomly assigned to the groups; (b) intact groups are used, but treatments are randomly assigned; and (c) intact groups are used where treatments occur naturally rather than being randomly assigned by the researcher. They maintain that ANCOVA is appropriate for the first situation, can be used with caution in the second, but should be abandoned altogether (perhaps in favor of the less restraining factorial block ANOVA) in the third. Two reasons are provided for their recommendations: first, it is never clear whether the covariance adjustment has removed all of the bias when proper randomization has not taken place, and second, when there are real differences among groups, covariance adjustments may involve computational extrapolation.

A number of researchers (e.g., Loftin and Madison, 1991; McLean, 1974; Raajimakers and Pieters, 1987; and Thompson, 1992) have addressed the issue of an unreliable covariate. Raajimakers and Pieters noted that there are two ways that a researcher can conceptualize covariate reliability. If one assumes the dependent is linearly related to the observed value of the covariate, then ANCOVA retains statistical validity. If, on the other hand, it is assumed that the dependent is linearly related to the underlying true score on the covariate (rather than the sample that was actually observed), then the F ratio will produce biased results. McLean's research, however, suggest that the issue of perfect reliability is less of a threat to the reliability of the F ratio if there is an independence of the covariate measure across treatment groups.

The Covariate's Relationship with the Independent and Dependent Variables

The covariate should be highly correlated with the dependent variable, yet should not be correlated with the independent. Feldt (1958) recommended use of a covariate only when the zero-order correlation between the covariate and dependent variable is $r \geq 0.6$. McLean (1979, 1989) saw the relationship between the covariate and independent variables to be the most fundamental of assumptions, suggesting that ANCOVA not be performed until after the data has been tested to see if it meets this assumption. If it is not met, the results are not invalidated as such, however it reduces ANCOVA's efficiency to slightly below that of ANOVA on the same data.

Homogeneity of Group Regression Slopes

This assumption requires that the slope of the regression line between the concomitant and dependent variables be the same for all levels of the grouping variable (see McLean, 1979, 1989; Thompson, 1992). The problem, if this assumption is violated, is analogous to interpreting main effects in the presence of significant interactions in an n-way factorial ANOVA. If heterogeneous slopes are suspect, the randomized block ANOVA is preferable to ANCOVA.

McClaren (1972) , Shields, (1978), and Johnson and Rakow (1993) empirically investigated the effects of violation of this assumption using unequal group sizes. These studies suggest that when the smallest regression coefficient and the largest variance are combined with

the smallest sample size, the empirical significance levels are biased in a non-conservative direction, and likewise, when the pairings are reversed, the test becomes conservative.

RESEARCH METHODOLOGY

This research explored the degree to which group sizes can differ before the robustness of analysis of variance (ANOVA) and analysis of covariance (ANCOVA) are jeopardized. Monte Carlo methodology was used, allowing for experimental investigation of potential threats to robustness under conditions common to researchers in education. Specifically, this research explored the effects of unequal group sizes with the following data set conditions: heterogeneity of group variances, skew, kurtosis, and (in ANCOVA) heterogeneity of regression slopes.

The Computing Environment and Programs Written to Conduct the Simulations

The statistical simulations were conducted on a VAX 6430 mainframe computer. A custom designed simulation package written in FORTRAN 77 was used, employing mathematical subroutines from the International Mathematical and Statistical Libraries (IMSL).

Two independent sets of simulations were conducted: one using a total group n of 90 and the other one using a total group n of 60. Experimentation was limited to simulations using three groups, and the total n was divided among the three groups in a systematic fashion designed to represent group size combinations similar to those faced by educational researchers.

The Simulation Process, Part I: Within a Single Replication of an Experiment

The data were created using the International Mathematical and Statistical Libraries (IMSL) subroutine RNVMN, which is designed to create multivariate nomal distributions with means of zero, variances of one, and correlations between vectors that can be specified by the user. Three vectors were created by IMSL: one dependent and two concomitant. Data for each treatment level were created separately and later merged, making it possible to obtain the unequal group slopes desired for the second concomitant vector. For the first concomitant, correlation between all groups and the IMSL created dependent vector was set at $r = .707$, thus simulating homogeneity of regression slopes. For the second concomitant, heterogeneity of slopes was simulated by creating concomitant vectors for group 1 that had a correlation of $r = .6$ with the group 1 dependent vector, a correlation of $r = .707$ between the second group concomitant and dependent vectors and $r = .8$ between the third group's concomitant and dependent vectors.

IMSL created dependent vectors which were normal around the first and second moments (i.e., "bell-shaped"). Since the integrity of the data generating process depended on these vectors being mesokurtic and with normal skew, a testing procedure was built into the

computer program to screen out vectors with skew and kurtosis that fall outside the 95% confidence intervals surrounding zero skew and kurtosis. Once an acceptable dependent vector was created, six additional copies were made and then perturbed so as to create vectors with the following shapes: platykurtic (skew = 0, kurtosis = -.5), leptokurtic (skew = 0, kurtosis = 2), moderate skew (skew = .5, kurtosis = 0), moderate skew and platykurtic (skew = .5, kurtosis = -.5), moderate skew and leptokurtc (skew = .5, kurtosis = 2), and extremely skewed and leptokurtic (skew = 1, kurtosis = 2). After perturbation, three duplicate copies of each of these seven vectors were created, and linearly transformed to create vectors with one of four degrees of variation: homogeneity of group variance (1:1:1), slight heterogeneity of variance (1:1/2:2), moderate heterogeneity of variance (1:2:3), and extreme heterogeneity of variance (1:3:5).

In the end 28 different dependent vectors, two concomitants and a grouping vector were created. For the ANOVA simulations, the grouping vector was combined with each dependent vector, resulting in 28 F ratios - one for each combination of skew, kurtosis, and variance. For ANCOVA, the first concomitant (with equal slopes) was combined with each of the 28 dependent vectors, then the second concomitant (with unequal slopes) was combined with each of the 28 dependent vectors. This produced an additional 56 ANCOVA F ratios.

In addition to using IMSL subroutines to generate the data, IMSL was also used to calculate the F ratios. Specifically, IMSL subroutine AONEW was used to obtain the ANOVA F values, while AONEC was used to calculate the ANCOVA F values.

The Simulation Process, Part II: The Global Design

As has been mentioned previously, two sets of analyses were run: one set containing simulations of three groups with a total n of 90 and the other set using simulations of three groups with a total n of 60. These sample sizes were chosen because they are similar to what educational researchers often encounter in practice. The use of two independent sets of simulations was done to allow for the cross-validation of data patterns that might be identified by the research.

For each of the two sets of analyses, the number assigned to each of the three groups was allowed to vary - thereby producing degrees of difference in the unequal n designs. When the total n was 90, the following combinations were tested: 30 in all three groups; groups of 28, 30 and 32; groups of 26, 30 and 34; groups of 24, 30 and 36; groups of 22, 30 and 38; and groups of 20, 30 and 40. These group size combinations resulted in the following percentages of difference between the largest and smallest group sizes: 0%, 12.5%, 24%, 33%, 42%, and 50%. When the total n was 60, the following combinations were tested: 20 in all three groups; groups of 18, 20 and 22; groups of 16, 20 and 24; groups of 14, 20 and 26; and groups of 12, 20 and 28. These group sizes resulted in the following differences: 0%, 18%, 33%, 46%, and 57%.

Each of these group size combinations were combined with one or more of the following data set violations: skew, kurtosis, heterogeneity of group variances and (in ANCOVA) homogeneity of regression slopes. This allowed for the identification of factors that threaten the robustness of ANOVA and ANCOVA, and allowed the determination when each of these factors should become a source of concern for the research practitioner.

Glass et al. (1972) recommended that sampling distributions created in Monte Carlo studies have a minimum of 2000 F ratios each. In those simulations involving homogeneity of group variances (1:1:1), the number of data points in the sampling distribution were 4000 - twice the recommendation of Glass et al. Where unequal n's were combined with heterogeneity of variances, sampling distributions of 2000 F ratios were created; each representing one of two experimental conditions which previous literature (e. g., Box, 1954; Glass, 1972) suggest will cause differing effects on type I error rates. For ANOVA, condition A paired the smallest n with the smallest variance, while the largest n was paired with the largest variance. Condition B paired the smallest n with the largest variance, while the largest n was paired with the smallest variance. For ANCOVA, condition A occured when the smallest n was joined with the smallest regression coefficient and the smallest variance, while the largest n was joined with the largest regression coefficient and the largest variance. Similarly, condition B occured when the smallest n was joined with the smallest regression coefficient and the largest variance, while the largest n was joined with the largest regression coefficient and the smallest variance.

After all simulations were run, 462 empirical sampling distributions of 4000 F ratios each were created, representing all single and compound violations for ANOVA and ANCOVA when homogeneity of variance was present. Another 2772 sampling distributions of 2000 F ratios each included all ANOVA and ANCOVA simulations with heterogeneity of variances present.

Of these 3234 F sampling distributions, eleven ANOVA and eleven ANCOVA F sampling distributions (one for each sample size combination) were created using data having no violation under study. These twenty-two sampling distributions served as a baseline for comparisons as well as also providing a check to make sure the simulation was operating properly.

Statistical Analysis of the Sampling Distributions

Data were analyzed using three different methods: (1) significance testing of sampling distribution tails (the area of the curve where hypothesis testing decisions are made), (2), through the comparison of nominal alpha levels with the actual (observed) significance levels found in the simulation process, and (3) graphically. Significance testing was performed using the Kolmogorov-Smirnov one sample test at the $\alpha < .05$ and (where applicable) $\alpha < .01$ levels of significance. This procedure allows comparison of an empirical data distribution with a standard distribution. Specific to this research, the tail regions of each of the 3234 F sampling distributions

were compared against the tail region of the appropriate theoretical F sampling distribution. Significant differences suggest a lack of robustness for the single or compound violation tested.

In addition to statistical testing, comparisons of the actual significance levels found in the simulations were compared against the nominal statistical level (that is, the statistical level a researcher believes he or she is using). The three nominal statistical levels examined in this research are $\alpha < .10$, $\alpha < .05$ and $\alpha < .01$. It is customary for a researcher to choose one of these significance levels in his or her own hypotheses testing - accepting at faith the fact that it represents the true probability of making type I error. The actual significance level, on the other hand, represents a more accurate estimate of the probability of making type I error when faced with a specific set of mathematical conditions.

RESULTS

Summarized below are the effects of violations of data set assumptions for the ANOVA and ANCOVA simulations included in this study. Since the integrity of the results is dependent on the quality of the data generated, the first section will discuss the population data. The second section will discuss simulations with homogeneity of variances. The next two sections will summarize the results of ANOVA and ANCOVA simulations with heterogeneity of variances. Finally, a summary of the results of the graphic analysis is included.

The results discussed below are analyzed using three different methodologies. The first method uses the Kolmogorov-Smirnov test on each of the 3234 sampling distributions. The KS test compares the tail region of each empirical sampling distribution against the appropriate theoretical (i.e., nominal) F distribution to determine if the empirical distribution tail is significantly different from theoretical F. The second method compares the observed significance levels (i.e., the observed type I error rate) for each of the 3234 mathematical conditions against the nominal alpha level (i.e., the expected type I error rate). To conserve space, separate KS tables are not included here. Instead significance testing results are color coded into the actual significance level tables. Finally, the third method involves examinination of the data graphs.

Analysis of the Population Data

Data created in each replication were retained in order to verify the integrity of the results. Vectors for each treatment level were created individually. Later they were merged with vectors for the other treatment levels so that ANOVA and ANCOVA could be performed. Population vectors were checked twice: once for each treatment level separately, and again after treatment levels were merged. All population vectors, including both those created originally by IMSL and those that were the result of algebraic transformation, were at or near the target parameters.

Population sizes are worth noting. In their classic 1972 paper, Glass et al. suggested that populations with the desired characteristics have a minimum 10,000 data points each. Population N's used in this study were considerably larger than that minimum: for the full (merged) vectors having homogeneity of variance, the N's were 360,000 when the total experiment n was 90 and 240,000 when the total n was 90. For full (merged) vectors having heterogeneity of variance, the N's were 180,000 when n was 90 and 120,000 when n was 60.

Effects of Data Set Violations When Homogeneity of Variance is Present

Results When the Total n = 90

In the absence of heterogeneity of group variances, no significant differences emerged between the nominal F distribution and the empirical F sampling distributions until the magnitude of group size differences became 50% (the 20:30:40 simulation). When group n's became this extreme, KS testing revealed that all ANCOVA simulations having unequal group regression slopes were significantly different from the theoretical F distribution at the $\alpha < .01$ level.

With a nominal significance (i.e., alpha) level of $\alpha < .10$, a researcher expects 10% chance of making type I error. Likewise, at nominal significance levels of $\alpha < .05$ and $\alpha < .01$, the research practitioner expects 5% and 1% chance of type I error respectively. When homogeneity of variance is present, the actual (i.e., observed) differences for ANOVA simulations as well as all equal slope ANCOVA simulations were found to be within ±.01 of the expected levels at $\alpha < .10$ and $\alpha < .05$, and within ± .003 at $\alpha < .01$. This was true even when the difference between the largest and smallest group size was as large as 50%.

When ANCOVA involved unequal regression slopes, on the other hand, larger differences from theoretical alpha emerged when group n's were very unequal. For the 22:30:38 simulation (where the groups differed by 42%), the observed risk of type I error was found to be between 9% and 13% (i.e., ±.02 of nominal alpha) for the nominal level of $\alpha < .10$. When the group sizes were 20:30:40, an even larger discrepancy emerged at the $\alpha < .10$ level (±.04 of nominal alpha). Observed error rates for $\alpha < .05$ fell between .05 and .07, while for $\alpha < .01$ they were between .008 and .022 (see table 1).

Results When the Total n = 60

In the absence of heterogeneity of group variances, no significant differences emerged between the nominal and empirical F distributions until the group size differences became 57% (the 12:20:28 simulation). When group n's were this extreme however, KS testing revealed that leptokurtic ANCOVA simulations having unequal slopes were significant at the $\alpha < .05$ level.

At the nominal level of $\alpha < .10$, the observed significance levels for ANOVA and those ANCOVA's with equal slopes varied ±.01 from nominal alpha , no matter what the degree of

Table 1: Actual Significance Levels when Homogeneity of Variance is Present

Group Size Combinations	30 : 30 : 30			28 : 30 : 32			26 : 30 : 34			24 : 30 : 36			22 : 30 : 38			20 : 30 : 40		
Nominal Alpha Levels	p≤.10	p≤.05	p≤.01	p≤.10	p≤.05	p≤.01	p≤.10	p≤.05	p≤.01	p≤.10	p≤.05	p≤.01	p≤.10	p≤.05	p≤.01	p≤.10	p≤.05	p≤.01
ANOVA																		
Distributional Shape																		
Normal	.10	.05	.009	.11	.05	.012	.09	.05	.008	.10	.05	.014	.10	.05	.009	.10	.05	.013
Platykurtic	.10	.05	.009	.10	.05	.012	.09	.05	.008	.10	.05	.014	.09	.05	.008	.10	.05	.013
Leptokurtic	.10	.05	.012	.11	.06	.012	.10	.04	.008	.10	.05	.013	.10	.05	.009	.11	.05	.012
Moderate Skew	.10	.05	.009	.10	.05	.012	.10	.04	.008	.10	.05	.015	.10	.04	.010	.10	.05	.013
Mod. Skew & Platy	.09	.04	.008	.10	.05	.012	.09	.05	.008	.10	.05	.014	.09	.04	.009	.09	.04	.013
Mod. Skew & Lepto	.10	.05	.008	.11	.05	.012	.10	.05	.010	.10	.05	.013	.10	.05	.010	.11	.05	.012
Ext. Skew & Lepto	.10	.04	.009	.10	.06	.011	.10	.05	.010	.10	.05	.014	.10	.05	.011	.10	.05	.012
ANCOVA: WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.10	.05	.009	.10	.05	.013	.10	.05	.010	.10	.05	.011	.09	.05	.009	.10	.05	.013
Platykurtic	.10	.05	.009	.10	.05	.013	.10	.05	.010	.10	.05	.012	.09	.05	.009	.10	.05	.013
Leptokurtic	.10	.06	.012	.10	.05	.012	.10	.05	.010	.10	.05	.011	.10	.05	.009	.11	.05	.012
Moderate Skew	.10	.05	.009	.10	.05	.013	.10	.05	.010	.11	.05	.010	.09	.05	.007	.10	.05	.013
Mod. Skew & Platy	.09	.05	.009	.10	.05	.013	.10	.05	.010	.10	.05	.009	.09	.04	.006	.09	.04	.013
Mod. Skew & Lepto	.10	.06	.008	.10	.05	.012	.10	.05	.010	.10	.05	.011	.09	.04	.009	.11	.05	.012
Ext. Skew & Lepto	.10	.06	.009	.10	.05	.012	.10	.05	.010	.10	.05	.010	.09	.04	.008	.10	.05	.012
ANCOVA : WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.09	.05	.008	.10	.05	.013	.10	.05	.012	.12	.06	.022	.12	.07	.016	.13	.07	.017
Platykurtic	.09	.05	.008	.10	.05	.013	.10	.05	.013	.11	.06	.022	.13	.07	.016	.14	.07	.016
Leptokurtic	.10	.05	.012	.11	.05	.012	.11	.05	.012	.12	.06	.021	.13	.07	.016	.14	.07	.015
Moderate Skew	.10	.05	.009	.10	.05	.013	.10	.05	.013	.12	.06	.017	.13	.07	.016	.14	.07	.016
Mod. Skew & Platy	.10	.05	.008	.10	.05	.013	.10	.05	.012	.11	.06	.017	.13	.07	.016	.14	.07	.015
Mod. Skew & Lepto	.10	.05	.008	.11	.05	.012	.10	.05	.013	.11	.06	.018	.13	.07	.020	.14	.07	.016
Ext. Skew & Lepto	.10	.05	.009	.10	.05	.012	.09	.05	.014	.10	.05	.017	.13	.07	.015	.14	.07	.015

Group Size Combinations	20 : 20 : 20			18 : 20 : 22			16 : 20 : 24			14 : 20 : 26			12 : 20 : 28		
ANOVA															
Distributional Shape															
Normal	.10	.05	.007	.10	.05	.010	.10	.05	.009	.09	.04	.009	.09	.05	.008
Platykurtic	.10	.04	.007	.10	.05	.010	.10	.05	.009	.09	.04	.010	.09	.04	.008
Leptokurtic	.09	.05	.007	.10	.05	.010	.11	.05	.008	.10	.05	.011	.10	.05	.009
Moderate Skew	.09	.04	.007	.10	.05	.010	.10	.05	.010	.09	.04	.010	.09	.05	.009
Mod. Skew & Platy	.10	.05	.007	.09	.05	.010	.10	.05	.010	.09	.04	.010	.09	.05	.010
Mod. Skew & Lepto	.10	.05	.008	.10	.05	.010	.10	.05	.008	.10	.05	.011	.09	.05	.009
Ext. Skew & Lepto	.10	.04	.007	.10	.05	.010	.10	.05	.008	.09	.05	.010	.09	.05	.009
ANCOVA : WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.10	.05	.008	.09	.04	.011	.10	.05	.009	.10	.05	.009	.10	.05	.013
Platykurtic	.10	.05	.007	.09	.04	.011	.10	.05	.010	.09	.05	.010	.10	.05	.012
Leptokurtic	.10	.05	.007	.09	.05	.011	.10	.05	.011	.10	.05	.009	.10	.05	.013
Moderate Skew	.10	.05	.010	.09	.04	.011	.10	.05	.010	.09	.05	.010	.09	.05	.012
Mod. Skew & Platy	.10	.05	.009	.09	.04	.009	.09	.05	.010	.09	.04	.009	.09	.05	.012
Mod. Skew & Lepto	.10	.06	.008	.09	.04	.011	.10	.05	.011	.10	.05	.009	.10	.05	.013
Ext. Skew & Lepto	.10	.05	.005	.09	.04	.011	.10	.05	.010	.10	.05	.010	.10	.05	.013
ANCOVA : WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.09	.05	.009	.10	.05	.011	.11	.06	.011	.12	.06	.016	.12	.06	.014
Platykurtic	.09	.05	.010	.10	.05	.011	.11	.05	.010	.11	.06	.016	.12	.06	.013
Leptokurtic	.09	.05	.009	.11	.05	.011	.11	.05	.010	.12	.06	.016	.12	.06	.015
Moderate Skew	.09	.05	.010	.11	.05	.012	.11	.05	.012	.11	.06	.015	.12	.06	.014
Mod. Skew & Platy	.09	.05	.009	.10	.05	.011	.10	.05	.013	.11	.06	.013	.12	.06	.014
Mod. Skew & Lepto	.09	.05	.009	.11	.05	.010	.11	.05	.010	.12	.06	.015	.12	.06	.015
Ext. Skew & Lepto	.09	.05	.010	.11	.05	.011	.11	.05	.010	.12	.06	.013	.12	.06	.014

Empirical distribution found significantly different from the theoretical F distribution at the $p < .05$ level
Empirical distribution found significantly different from the theoretical F distribution at the $p < .01$ level

BEST COPY AVAILABLE

difference between the largest and smallest groups. Furthermore, even the ANCOVA analyses involving unequal slopes deviated from nominal alpha by only ±.02. At nominal alpha levels of $\alpha < .05$ and $\alpha < .01$, most observed alpha levels were within ±.01.

Effects of Data Set Assumptions When Heterogeneity of Variance is Present and Total n = 90

Effects of Heterogeneity of Variances in the Balanced Design : 30 : 30 : 30

When three groups of 30 are used, KS tests reveal no statistically significant differences between the tail regions of the empirical sampling distributions created in the simulation process and the theoretical F distribution - even in ANCOVA simulations when heterogeneity of variance (ratios of 1:3:5) is joined with unequal regression slopes. The observed significance levels are all either equal to or within ±.01 of the appropriate nominal alpha level. This, of course, suggests that in the balanced design the true risk of making type I error is equal to or very near the risk that the researcher believes it to be (see tables 2, 3 and 4).

Effects of Heterogeneity of Variances in the Design : 28 : 30 : 32

In these analyses the difference between the smallest and largest group sizes is 12.5%. With such close group sizes, KS testing reveals no significant differences from theoretical F. In most of the simulations, actual significance levels are within ±.01 of the intended significance level. However, observed significance levels are off by +.02 in some simulations - primarily condition B ANCOVA simulations having extreme heterogeneity of variances (one difference is +.03). Condition B ANCOVA analyses, as one will recall, are those where the smallest group size and slope are paired with the largest group variance, while condition A ANCOVA analyses matched the smallest group size and slope with the smallest variance(see tables 2, 3 and 4).

Effects of Heterogeneity of Variances in the Design : 26 : 30 : 34

In these analyses the difference between the largest and smallest group is 24%. With group differences of this magnitude, statistically significant differences between the empirical and theoretical sampling distributions begin to emerge- although they are confined to Condition B ANCOVAs when both extreme heterogeneity of variances and unequal slopes are present.

For condition A at the nominal $\alpha < .10$ level of significance, the observed significance levels range between .08 and .10. At nominal level of $\alpha < .05$, most observed levels are between .04 and .06. At the nominal level of $\alpha < .01$, the observed levels fall between .008 and .013.

More variation is found among condition B simulations, where at the nominal level of $\alpha < .10$, observed significance levels fall between .10 and .12. At the nominal level of $\alpha < .05$, observed significance level fall between .06 and .07, while at the nominal level of $\alpha < .01$ the actual significance levels fall between .017 and .029 (see tables 2, 3 and 4).

Table 2: Actual Significance Levels When Using the Nominal Significance of $p < .10$ and Total $n = 90$

Group Size Condition A	30 : 30 : 30			28 : 30 : 32			26 : 30 : 34			24 : 30 : 36			22 : 30 : 38			20 : 30 : 40		
Group Size Condition B	30 : 30 : 30			32 : 30 : 28			34 : 30 : 26			36 : 30 : 24			38 : 30 : 22			40 : 30 : 20		
Largest to Smallest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A																		
Distributional Shape																		
Normal	.09	.10	.10	.10	.10	.10	.09	.09	.09	.08	.08	.07	.08	.08	.07	.09	.08	.07
Platykurtic	.09	.09	.10	.10	.10	.10	.08	.08	.09	.08	.08	.08	.08	.07	.07	.08	.07	.07
Leptokurtic	.09	.10	.10	.10	.10	.11	.08	.08	.09	.08	.07	.07	.09	.08	.07	.08	.06	.07
Moderate Skew	.09	.10	.10	.10	.10	.10	.09	.09	.09	.08	.08	.07	.08	.07	.07	.08	.07	.07
Mod. Skew & Platy	.09	.09	.10	.10	.10	.10	.09	.08	.08	.08	.08	.08	.08	.08	.07	.08	.07	.07
Mod. Skew & Lepto	.10	.10	.10	.10	.10	.10	.10	.09	.09	.08	.07	.07	.09	.08	.07	.09	.08	.07
Ext. Skew & Lepto	.10	.10	.10	.10	.10	.10	.09	.09	.09	.08	.08	.07	.09	.08	.07	.09	.08	.07
ANOVA : CONDITION B																		
Distributional Shape																		
Normal	.10	.10	.10	.11	.11	.11	.10	.11	.11	.12	.13	.14	.12	.13	.15	.12	.15	.17
Platykurtic	.10	.10	.10	.11	.11	.12	.10	.11	.11	.12	.13	.14	.12	.14	.14	.12	.15	.17
Leptokurtic	.10	.10	.11	.11	.11	.12	.10	.11	.12	.12	.13	.15	.12	.14	.16	.13	.15	.17
Moderate Skew	.10	.10	.10	.11	.11	.12	.10	.11	.11	.12	.13	.14	.12	.13	.14	.13	.15	.17
Mod. Skew & Platy	.09	.09	.10	.10	.11	.11	.10	.10	.11	.12	.13	.14	.12	.13	.14	.13	.15	.17
Mod. Skew & Lepto	.10	.11	.10	.11	.11	.12	.11	.11	.12	.13	.14	.15	.13	.14	.15	.13	.15	.17
Ext. Skew & Lepto	.10	.10	.11	.11	.11	.12	.11	.12	.12	.12	.13	.14	.12	.14	.15	.14	.15	.17
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.10	.10	.10	.09	.09	.09	.09	.09	.08	.08	.08	.08	.07	.07	.06	.07	.07	.06
Platykurtic	.10	.10	.10	.09	.09	.09	.09	.09	.08	.08	.08	.08	.07	.07	.06	.07	.06	.06
Leptokurtic	.10	.11	.11	.09	.09	.09	.09	.09	.08	.08	.08	.08	.07	.06	.06	.07	.07	.06
Moderate Skew	.09	.10	.11	.09	.09	.09	.09	.09	.08	.08	.08	.08	.07	.06	.06	.07	.07	.06
Mod. Skew & Platy	.09	.10	.10	.09	.09	.09	.09	.09	.08	.09	.08	.08	.07	.06	.06	.07	.06	.06
Mod. Skew & Lepto	.10	.10	.11	.10	.09	.09	.09	.08	.08	.08	.08	.08	.07	.07	.07	.07	.07	.07
Ext. Skew & Lepto	.10	.10	.11	.09	.09	.10	.09	.08	.08	.08	.08	.08	.07	.07	.06	.07	.07	.06
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.10	.10	.10	.10	.11	.12	.11	.11	.12	.13	.13	.13	.13	.14	.15	.14	.16	.17
Platykurtic	.10	.10	.10	.10	.11	.11	.11	.11	.12	.12	.13	.14	.12	.14	.15	.14	.16	.17
Leptokurtic	.10	.10	.10	.10	.11	.12	.12	.12	.12	.12	.13	.14	.13	.14	.15	.14	.16	.17
Moderate Skew	.11	.11	.10	.11	.11	.13	.12	.12	.13	.13	.14	.14	.13	.14	.15	.14	.16	.18
Mod. Skew & Platy	.10	.10	.10	.10	.11	.12	.12	.13	.13	.12	.13	.14	.12	.14	.15	.14	.16	.18
Mod. Skew & Lepto	.10	.10	.11	.11	.11	.12	.10	.11	.12	.12	.13	.14	.12	.14	.15	.14	.16	.17
Ext. Skew & Lepto	.10	.10	.11	.11	.11	.12	.10	.11	.12	.12	.13	.14	.13	.14	.15	.14	.17	.18
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.09	.09	.10	.11	.10	.11	.09	.09	.08	.09	.08	.08	.10	.09	.08	.10	.09	.08
Platykurtic	.09	.09	.09	.11	.10	.11	.09	.08	.08	.09	.08	.08	.09	.08	.08	.10	.09	.08
Leptokurtic	.09	.10	.10	.11	.11	.10	.09	.09	.09	.09	.08	.08	.10	.09	.09	.10	.09	.08
Moderate Skew	.10	.09	.10	.10	.10	.10	.09	.09	.08	.08	.08	.08	.09	.08	.07	.10	.09	.08
Mod. Skew & Platy	.10	.10	.09	.10	.10	.10	.09	.08	.08	.08	.08	.07	.09	.08	.07	.10	.09	.08
Mod. Skew & Lepto	.09	.10	.10	.11	.11	.10	.09	.08	.08	.08	.08	.08	.10	.09	.08	.09	.09	.08
Ext. Skew & Lepto	.09	.09	.10	.11	.11	.11	.09	.08	.08	.09	.08	.08	.10	.09	.08	.10	.09	.08
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.10	.10	.10	.11	.12	.12	.12	.13	.13	.13	.14	.15	.14	.15	.16	.17	.18	.19
Platykurtic	.09	.10	.10	.11	.12	.12	.12	.13	.13	.13	.14	.15	.14	.15	.16	.17	.18	.20
Leptokurtic	.10	.10	.10	.11	.12	.12	.12	.12	.13	.13	.14	.15	.14	.15	.16	.17	.18	.20
Moderate Skew	.10	.10	.10	.10	.11	.12	.12	.13	.13	.13	.14	.15	.15	.16	.17	.17	.18	.20
Mod. Skew & Platy	.10	.10	.10	.10	.11	.12	.12	.13	.13	.13	.14	.15	.14	.15	.16	.16	.18	.20
Mod. Skew & Lepto	.11	.11	.11	.11	.12	.12	.11	.11	.11	.13	.14	.15	.15	.16	.17	.17	.19	.20
Ext. Skew & Lepto	.11	.11	.11	.11	.11	.12	.10	.11	.12	.13	.14	.15	.15	.16	.18	.17	.19	.20

Empirical distribution found significantly different from the theoretical F distribution at the $p < .05$ level
Empirical distribution found significantly different from the theoretical F distribution at the $p < .01$ level

Table 3: Actual Significance Levels When Using the Nominal Significance of p < .05 and Total n = 90

Group Sizes Condition A	30 : 30 : 30			28 : 30 : 32			26 : 30 : 34			24 : 30 : 36			22 : 30 : 38			20 : 30 : 40		
Group Sizes Condition B	30 : 30 : 30			32 : 30 : 28			34 : 30 : 26			36 : 30 : 24			38 : 30 : 22			40 : 30 : 20		
Largest to Smallest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A																		
Distributional Shape																		
Normal	.05	.05	.05	.06	.06	.06	.05	.05	.05	.04	.04	.04	.03	.03	.03	.04	.03	.03
Platykurtic	.05	.05	.05	.06	.06	.06	.05	.05	.05	.05	.04	.04	.04	.03	.03	.03	.03	.03
Leptokurtic	.05	.05	.06	.06	.06	.06	.05	.05	.05	.04	.04	.04	.03	.03	.03	.04	.03	.03
Moderate Skew	.05	.05	.05	.05	.05	.06	.04	.05	.05	.05	.04	.04	.03	.03	.03	.03	.03	.03
Mod. Skew &Platy	.04	.05	.05	.05	.05	.06	.04	.05	.05	.04	.04	.04	.04	.03	.03	.03	.03	.03
Mod. Skew & Lepto	.05	.05	.05	.06	.06	.06	.05	.05	.05	.04	.04	.04	.03	.03	.03	.04	.03	.03
Ext. Skew & Lepto	.04	.05	.05	.06	.06	.06	.05	.05	.05	.04	.04	.04	.04	.03	.03	.03	.03	.03
ANOVA : CONDITION B																		
Distributional Shape																		
Normal	.05	.06	.06	.06	.06	.07	.06	.06	.07	.07	.07	.09	.06	.07	.09	.07	.08	.10
Platykurtic	.05	.05	.06	.06	.06	.06	.06	.06	.07	.06	.07	.08	.06	.07	.09	.07	.08	.10
Leptokurtic	.06	.05	.06	.06	.07	.07	.06	.06	.07	.07	.07	.08	.06	.07	.09	.07	.09	.10
Moderate Skew	.05	.06	.06	.06	.06	.06	.06	.06	.07	.06	.07	.08	.06	.08	.09	.07	.09	.10
Mod. Skew & Platy	.05	.05	.06	.05	.06	.06	.06	.06	.07	.06	.07	.08	.06	.08	.09	.07	.09	.10
Mod. Skew & Lepto	.05	.06	.08	.06	.06	.07	.06	.06	.07	.07	.08	.08	.06	.08	.09	.07	.09	.11
Ext. Skew & Lepto	.05	.06	.06	.06	.06	.07	.06	.06	.07	.07	.07	.08	.06	.08	.09	.07	.09	.11
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.05	.06	.06	.05	.05	.05	.04	.04	.04	.04	.04	.04	.03	.03	.03	.03	.03	.03
Platykurtic	.05	.06	.06	.05	.05	.05	.04	.04	.04	.04	.04	.04	.03	.03	.03	.03	.03	.03
Leptokurtic	.05	.05	.06	.05	.05	.05	.04	.04	.04	.04	.04	.04	.03	.03	.03	.03	.03	.03
Moderate Skew	.06	.06	.06	.05	.05	.05	.04	.04	.05	.04	.04	.04	.03	.03	.03	.03	.03	.03
Mod. Skew& Platy	.05	.06	.05	.05	.05	.05	.04	.04	.05	.04	.04	.04	.03	.03	.02	.03	.03	.03
Mod. Skew & Lepto	.05	.05	.06	.05	.04	.05	.04	.04	.04	.04	.04	.04	.03	.03	.03	.03	.03	.03
Ext. Skew & Lepto	.06	.06	.06	.05	.05	.05	.04	.04	.04	.04	.04	.04	.03	.03	.03	.03	.03	.03
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.05	.05	.05	.06	.06	.07	.06	.07	.08	.07	.07	.09	.06	.08	.09	.07	.09	.11
Platykurtic	.05	.05	.05	.06	.06	.06	.06	.07	.08	.07	.08	.08	.06	.07	.09	.07	.09	.11
Leptokurtic	.06	.06	.06	.06	.06	.07	.06	.06	.07	.06	.07	.08	.07	.08	.09	.07	.09	.11
Moderate Skew	.05	.05	.05	.06	.06	.06	.06	.07	.07	.07	.07	.08	.06	.07	.09	.07	.09	.10
Mod. Skew & Platy	.05	.05	.06	.06	.06	.06	.06	.06	.07	.07	.08	.08	.06	.07	.09	.08	.08	.10
Mod. Skew & Lepto	.06	.06	.06	.06	.06	.06	.05	.06	.07	.07	.07	.08	.06	.07	.09	.08	.09	.11
Ext. Skew & Lepto	.06	.06	.06	.06	.06	.07	.06	.06	.07	.07	.08	.08	.06	.07	.09	.08	.09	.11
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.05	.05	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
Platykurtic	.04	.05	.05	.05	.06	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
Leptokurtic	.04	.05	.05	.05	.05	.05	.05	.05	.05	.04	.04	.04	.05	.04	.04	.05	.04	.04
Moderate Skew	.04	.05	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
Mod. Skew & Platy	.04	.05	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
Mod. Skew & Lepto	.04	.05	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
Ext. Skew & Lepto	.04	.05	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05	.04	.04
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.05	.05	.05	.06	.06	.07	.07	.07	.08	.07	.08	.09	.08	.09	.10	.10	.12	.13
Platykurtic	.05	.05	.05	.06	.06	.07	.07	.07	.08	.07	.09	.09	.08	.09	.10	.10	.12	.13
Leptokurtic	.05	.05	.05	.06	.06	.07	.07	.07	.08	.07	.08	.08	.08	.09	.10	.10	.12	.13
Moderate Skew	.05	.05	.06	.06	.06	.06	.07	.07	.08	.07	.08	.09	.08	.10	.11	.10	.12	.13
Mod. Skew & Platy	.05	.06	.06	.07	.07	.07	.07	.07	.08	.07	.08	.08	.08	.09	.10	.10	.11	.13
Mod. Skew & Lepto	.05	.05	.06	.06	.06	.07	.06	.06	.07	.07	.08	.08	.08	.09	.10	.11	.12	.13
Ext. Skew & Lepto	.05	.06	.06	.06	.06	.07	.06	.06	.07	.07	.08	.09	.08	.09	.11	.10	.11	.13

Empirical distribution found significantly different from the theoretical F distribution at the p < .05 level
Empirical distribution found significantly different from the theoretical F distribution at the p < .01 level

Table 4: Actual Significance Levels When Using the Nominal Significance of p < .01 and Total n = 90

Group Sizes Condition A	30 : 30 : 30			28 : 30 : 32			26 : 30 : 34			24 : 30 : 36			22 : 30 : 38			20 : 30 : 40		
Group Sizes Condition B	30 : 30 : 30			32 : 30 : 28			34 : 30 : 26			36 : 30 : 24			38 : 30 : 22			40 : 30 : 20		
Largest to Smallest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A																		
Distributional Shape																		
Normal	.009	.009	.012	.013	.014	.013	.009	.010	.011	.008	.007	.008	.006	.005	.006	.008	.007	.007
Platykurtic	.008	.009	.013	.012	.012	.013	.009	.010	.010	.009	.007	.008	.005	.005	.005	.009	.007	.007
Leptokurtic	.009	.009	.011	.014	.015	.016	.009	.011	.011	.008	.008	.008	.008	.007	.007	.007	.006	.006
Moderate Skew	.008	.010	.011	.014	.015	.017	.009	.009	.011	.009	.008	.009	.006	.006	.006	.008	.008	.008
Mod. Skew & Platy	.009	.009	.009	.013	.016	.016	.008	.009	.011	.007	.007	.009	.006	.005	.005	.009	.008	.007
Mod. Skew & Lepto	.009	.009	.010	.016	.015	.016	.011	.012	.012	.007	.008	.010	.009	.007	.007	.009	.007	.006
Ext. Skew & Lepto	.008	.009	.011	.014	.014	.016	.012	.013	.013	.008	.009	.009	.007	.006	.007	.009	.008	.009
ANOVA : CONDITION B																		
Distributional Shape																		
Normal	.010	.011	.012	.012	.016	.020	.015	.019	.021	.022	.025	.031	.014	.019	.027	.022	.029	.037
Platykurtic	.010	.011	.012	.012	.014	.020	.016	.017	.021	.022	.028	.030	.013	.018	.028	.022	.029	.034
Leptokurtic	.010	.010	.012	.013	.015	.016	.014	.017	.020	.021	.023	.028	.015	.017	.028	.023	.029	.037
Moderate Skew	.009	.010	.013	.014	.017	.020	.015	.019	.025	.023	.028	.032	.014	.020	.027	.023	.030	.039
Mod. Skew & Platy	.009	.011	.013	.014	.018	.020	.017	.019	.023	.023	.029	.032	.014	.021	.028	.023	.029	.036
Mod. Skew & Lepto	.009	.010	.012	.013	.015	.019	.012	.013	.019	.021	.025	.029	.014	.019	.023	.024	.027	.035
Ext. Skew & Lepto	.010	.010	.013	.013	.014	.019	.011	.013	.017	.024	.026	.029	.015	.019	.024	.023	.027	.034
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.017	.018	.018	.008	.008	.011	.011	.011	.011	.010	.009	.009	.009	.008	.008	.010	.008	.008
Platykurtic	.017	.016	.018	.007	.009	.011	.010	.010	.012	.009	.009	.010	.009	.008	.008	.010	.008	.007
Leptokurtic	.017	.017	.017	.007	.009	.010	.010	.008	.009	.008	.008	.008	.008	.007	.008	.009	.009	.006
Moderate Skew	.016	.016	.016	.008	.008	.011	.010	.011	.010	.011	.009	.008	.009	.007	.009	.010	.009	.008
Mod. Skew & Platy	.014	.015	.015	.004	.006	.010	.010	.011	.011	.010	.009	.009	.007	.008	.009	.009	.007	.006
Mod. Skew & Lepto	.019	.020	.017	.008	.009	.010	.011	.011	.008	.009	.009	.009	.008	.007	.009	.010	.008	.007
Ext. Skew & Lepto	.017	.019	.018	.008	.009	.010	.012	.011	.010	.010	.011	.010	.006	.007	.008	.008	.008	.007
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.010	.011	.011	.013	.017	.020	.013	.016	.020	.021	.021	.026	.016	.020	.024	.019	.028	.034
Platykurtic	.010	.010	.009	.014	.018	.020	.012	.015	.019	.021	.022	.027	.016	.019	.022	.021	.027	.033
Leptokurtic	.010	.012	.012	.013	.013	.017	.013	.017	.018	.020	.023	.025	.015	.019	.023	.017	.027	.034
Moderate Skew	.012	.009	.009	.014	.015	.017	.012	.015	.020	.019	.020	.027	.014	.020	.022	.015	.028	.037
Mod. Skew & Platy	.010	.011	.011	.014	.017	.018	.012	.016	.021	.016	.021	.027	.014	.019	.021	.014	.022	.034
Mod. Skew & Lepto	.010	.012	.012	.013	.014	.017	.015	.017	.021	.019	.020	.025	.015	.021	.024	.017	.026	.034
Ext. Skew & Lepto	.011	.010	.012	.013	.013	.016	.013	.015	.020	.017	.019	.025	.015	.020	.026	.017	.022	.032
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.007	.005	.007	.010	.009	.011	.009	.010	.010	.010	.011	.009	.006	.007	.008	.004	.004	.004
Platykurtic	.007	.006	.007	.010	.011	.011	.010	.009	.009	.010	.010	.010	.006	.006	.006	.004	.004	.004
Leptokurtic	.008	.008	.008	.009	.011	.012	.010	.009	.009	.008	.011	.010	.007	.008	.009	.006	.004	.002
Moderate Skew	.007	.005	.005	.008	.010	.012	.011	.009	.010	.008	.009	.012	.005	.007	.007	.006	.005	.004
Mod. Skew & Platy	.005	.004	.005	.009	.009	.011	.013	.011	.010	.007	.008	.009	.004	.006	.007	.006	.006	.005
Mod. Skew & Lepto	.007	.007	.005	.010	.011	.012	.009	.009	.008	.011	.010	.009	.008	.008	.008	.005	.004	.002
Ext. Skew & Lepto	.006	.004	.005	.011	.012	.013	.010	.011	.011	.011	.010	.010	.008	.007	.006	.005	.004	.004
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES																		
Distributional Shape																		
Normal	.013	.012	.015	.019	.021	.022	.020	.022	.028	.027	.030	.033	.023	.028	.033	.028	.036	.046
Platykurtic	.011	.012	.016	.018	.020	.023	.019	.023	.026	.027	.027	.033	.021	.027	.032	.028	.036	.044
Leptokurtic	.012	.014	.015	.017	.018	.021	.017	.019	.025	.024	.030	.032	.020	.029	.033	.029	.037	.048
Moderate Skew	.012	.013	.016	.017	.018	.022	.018	.022	.029	.024	.029	.033	.021	.027	.033	.031	.038	.046
Mod. Skew & Platy	.012	.015	.015	.014	.018	.020	.017	.021	.026	.023	.028	.031	.022	.028	.034	.025	.032	.044
Mod. Skew & Lepto	.013	.012	.015	.018	.019	.021	.014	.019	.024	.025	.028	.032	.022	.028	.032	.028	.037	.047
Ext. Skew & Lepto	.012	.014	.016	.017	.019	.021	.013	.016	.021	.023	.027	.030	.022	.028	.035	.029	.037	.047

Empirical distribution found significantly different from the theoretical F distribution at the p < .05 level
Empirical distribution found significantly different from the theoretical F distribution at the p < .01 level

Effects of Heterogeneity of Variances in the Design : 24 : 30 : 36

In these analyses the differences between the smallest and largest group is 33%. With differences of this magnitude, more statistically significant deviations from theoretical F emerge. Among ANOVA designs, condition B is most likely to emerge as significant. All but one condition B simulation having moderate heterogeneity of variance emerge as significant at the $\alpha < .05$ level while simulations having extreme heterogeneity of variance emerge significant at the $\alpha < .01$ level. Among condition A ANOVA simulations, only three were significant at the $\alpha < .05$ level.

With ANCOVA, statistically significant differences from theoretical F are found only in condition B. There, nine equal slope simulations are significant at the $\alpha < .05$ level and another six at the $\alpha < .01$ level. When condition B is coupled with unequal slopes, all but one emerge statistically significant - most at the $\alpha < .01$ level.

At the nominal level of $\alpha < .10$, the observed significance levels for condition A ANOVA and ANCOVA analyses fall between .07 and .09. At the nominal level of $\alpha < .05$, observed levels of significance fall between .04 and .05, while at the nominal level of $\alpha < .01$ the observed significance levels are between .007 and .012. With Condition B simulations at the nominal $\alpha < .10$ level of significance, the observed significance levels range between .12 and .15. With the nominal level of $\alpha < .05$, observed levels range between .06 and .09. Finally, with the nominal level of $\alpha < .01$, observed significance levels are between .016 and .033 (see tables 2, 3 and 4).

Effects of Heterogeneity of Variances in the Design : 22 : 30 : 38

In these analyses, the difference between the smallest and largest group is 42%. Statistically significant differences from theoretical F emerge with condition A ANCOVA having equal regression slopes (most at the $\alpha < .05$ level), however only two condition A ANOVA simulations and none of the condition A ANCOVA's with unequal slopes are significantly different from the theoretical distribution. Differences are even more prevalent with condition B, where moderate and extreme heterogeneity of variance almost always emerge as significant at the $\alpha < .01$ level. Among Condition B ANCOVA's with unequal slopes, even simulations with slight heterogeneity of group variances are significant at the $\alpha < .01$ level.

For condition A simulations at the nominal level of $\alpha < .10$, observed significance levels are between .06 and .10, while they are between .02 and .05 for the nominal level of $\alpha < .05$ and between .004 and .009 for $\alpha < .01$. For condition B the ranges are between .12 and .18 for $\alpha < .10$, .06 and .11 for $\alpha < .05$ and between .013 and .035 for $\alpha < .01$ (see tables 2, 3 and 4).

Effects of Heterogeneity of Variances in the Design : 20 : 30 : 40

In this set of analyses the difference between smallest and largest group is 50%. Significant differences from theoretical F occur in almost all condition B simulations with even a

slight degree of heterogeneity of variance (most at the $\alpha < .01$ level). Condition A ANCOVA with equal slopes also produces many significant differences at least at the $\alpha < .05$ level when coupled with heterogeneity of variances. Interestingly, almost no condition A ANOVA simulations are significant, nor were any of the condition A ANCOVA simulations having unequal slopes.

For condition A simulations at the nominal level of $\alpha < .10$, the observed significance levels fall between .06 and .10, while they are between .03 and .05 for nominal $\alpha < .05$ and between .002 and .010 at the nominal level of $\alpha < .01$. For the nominal significance level of $\alpha < .10$, the observed levels varied between .12 and .20, while for the nominal level of $\alpha < .01$ the observed levels varied between .014 and .048 (see tables 2, 3 and 4).

Effects of Data Set Assumptions When Heterogeneity of Variance is Present and Total n = 60

Effects of Heterogeneity of Variances in the Balanced Design : 20 : 20 : 20

When three groups of 20 were used, KS testing revealed no significant differences between tail regions of the theoretical and empirical sampling distributions. Also observed significance levels were close to their nominal counterparts (usually within ±.01 of nominal alpha). Again this suggests that in the balanced designs the true risk of making type I error is equal to or near the risk chosen by the researcher (see tables 5, 6 and 7).

Effects of Heterogeneity of Variances in the Design : 18 : 20 : 22

In these analyses, the difference between smallest and largest groups is 18%. This difference is so small that no significant differences from theoretical F appear. When nominal alpha is $\alpha < .10$, the observed significance levels are within ±.01 for condition A ANOVA , condition A ANCOVA with unequal regression slopes and condition B ANCOVA with equal regression slopes. Meanwhile, the observed significance levels are within ±.02 for condition B ANOVA, and condition A ANCOVA with equal slopes. Finally, observed levels vary between .11 and .13 for condition B ANCOVA with unequal slopes. For the nominal alpha of $\alpha < .05$, all observed levels varied within ±.01 except for condition B ANOVA with extreme variances and condition B ANCOVA with unequal slopes where differences may be +.02 (see tables 5, 6 and 7).

Effects of Heterogeneity of Variances in the Design : 16 : 20 : 24

In these analyses, the difference between the smallest and largest group size is 33%. Significant differences from theoretical F begin to appear when group sizes vary to this degree. Significant differences at the $\alpha < .05$ level emerge in some condition A ANOVA's and condition A ANCOVA's with equal regression slopes and moderate or extreme heterogeneity of variances. Significant differences at either the $\alpha < .05$ or $\alpha < .01$ levels emerge in most condition B simulations having moderate or extreme heterogeneity of variances.

Table 5: Actual Significance Levels When Using the Nominal Significance of $p \leq .10$ and Total n = 60

Group Sizes Condition A	20 : 20 : 20			18 : 20 : 22			16 : 20 : 24			14 : 20 : 26			12 : 20 : 28		
Group Sizes Condition B	20 : 20 : 20			22 : 20 : 18			24 : 20 : 16			26 : 20 : 14			28 : 20 : 12		
Largest to Smallest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A															
Distributional Shape															
Normal	.09	.09	.09	.09	.09	.09	.08	.07	.07	.07	.06	.06	.07	.06	.05
Platykurtic	.09	.09	.09	.09	.09	.09	.08	.07	.07	.07	.06	.06	.07	.06	.05
Leptokurtic	.09	.09	.10	.10	.09	.09	.08	.08	.08	.07	.07	.06	.07	.06	.06
Moderate Skew	.09	.09	.10	.09	.09	.10	.08	.07	.07	.07	.06	.06	.07	.06	.05
Mod. Skew & Platy	.09	.09	.10	.09	.09	.09	.07	.07	.07	.06	.06	.06	.07	.06	.05
Mod. Skew & Lepto	.10	.09	.10	.10	.09	.09	.08	.08	.07	.08	.07	.06	.07	.06	.05
Ext. Skew & Lepto	.10	.10	.10	.10	.10	.10	.08	.08	.08	.08	.07	.06	.07	.06	.05
ANOVA : CONDITION B															
Distributional Shape															
Normal	.10	.10	.10	.10	.11	.11	.13	.14	.15	.12	.14	.15	.12	.15	.18
Platykurtic	.10	.10	.10	.10	.11	.11	.13	.14	.15	.12	.14	.16	.12	.15	.18
Leptokurtic	.10	.10	.11	.11	.11	.11	.13	.14	.15	.12	.14	.16	.13	.15	.17
Moderate Skew	.10	.10	.10	.10	.11	.12	.13	.13	.15	.12	.14	.16	.13	.15	.17
Mod. Skew & Platy	.09	.09	.10	.10	.11	.12	.13	.14	.15	.12	.14	.16	.13	.15	.18
Mod. Skew & Lepto	.10	.10	.10	.11	.11	.12	.13	.14	.15	.12	.14	.16	.13	.15	.18
Ext. Skew & Lepto	.10	.10	.10	.11	.11	.12	.12	.13	.15	.12	.14	.16	.13	.15	.18
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.10	.10	.10	.08	.08	.09	.08	.07	.07	.08	.07	.07	.08	.08	.07
Platykurtic	.10	.09	.10	.08	.08	.08	.07	.07	.07	.07	.07	.07	.08	.08	.07
Leptokurtic	.09	.10	.10	.08	.09	.08	.08	.07	.07	.08	.07	.07	.08	.07	.07
Moderate Skew	.09	.09	.10	.08	.08	.09	.08	.08	.07	.08	.08	.07	.08	.07	.07
Mod. Skew & Platy	.09	.09	.10	.08	.08	.08	.08	.08	.07	.08	.07	.07	.07	.07	.07
Mod. Skew & Lepto	.10	.10	.10	.08	.08	.09	.08	.08	.07	.08	.08	.07	.08	.08	.07
Ext. Skew & Lepto	.10	.10	.11	.08	.08	.09	.09	.09	.08	.08	.08	.07	.08	.07	.07
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.10	.10	.10	.10	.11	.11	.12	.13	.13	.12	.13	.15	.14	.15	.18
Platykurtic	.10	.10	.10	.10	.10	.11	.12	.13	.13	.12	.13	.15	.13	.15	.18
Leptokurtic	.10	.10	.10	.10	.11	.11	.12	.13	.14	.13	.13	.15	.14	.15	.17
Moderate Skew	.10	.10	.10	.10	.11	.11	.12	.13	.13	.13	.14	.15	.13	.15	.18
Mod. Skew & Platy	.10	.10	.10	.11	.11	.11	.12	.13	.14	.13	.15	.15	.13	.16	.18
Mod. Skew & Lepto	.10	.10	.10	.11	.11	.11	.13	.14	.14	.13	.14	.15	.13	.15	.17
Ext. Skew & Lepto	.10	.10	.10	.10	.11	.11	.13	.13	.14	.13	.14	.16	.13	.15	.17
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.09	.09	.09	.10	.09	.09	.09	.08	.08	.09	.09	.08	.09	.08	.07
Platykurtic	.09	.09	.09	.10	.09	.09	.09	.08	.07	.09	.09	.08	.09	.08	.07
Leptokurtic	.09	.09	.10	.10	.10	.09	.09	.08	.07	.10	.09	.09	.09	.08	.07
Moderate Skew	.09	.09	.09	.10	.09	.09	.08	.08	.08	.09	.09	.08	.09	.08	.07
Mod. Skew & Platy	.09	.09	.09	.09	.09	.09	.08	.07	.07	.09	.08	.07	.08	.07	.07
Mod. Skew & Lepto	.09	.09	.10	.10	.09	.10	.09	.08	.08	.10	.09	.09	.09	.08	.07
Ext. Skew & Lepto	.09	.09	.10	.10	.09	.09	.08	.08	.08	.10	.09	.08	.09	.08	.07
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.10	.10	.10	.12	.12	.12	.12	.14	.14	.15	.16	.17	.16	.18	.20
Platykurtic	.09	.09	.09	.12	.12	.12	.12	.13	.14	.15	.16	.17	.16	.18	.20
Leptokurtic	.09	.10	.10	.12	.12	.13	.13	.13	.15	.15	.16	.17	.16	.18	.20
Moderate Skew	.10	.10	.10	.12	.12	.12	.13	.13	.14	.16	.16	.18	.16	.18	.21
Mod. Skew & Platy	.09	.10	.09	.11	.11	.12	.12	.13	.14	.15	.17	.18	.16	.18	.20
Mod. Skew & Lepto	.10	.11	.10	.12	.12	.13	.13	.14	.14	.16	.17	.18	.16	.19	.20
Ext. Skew & Lepto	.11	.10	.11	.12	.12	.13	.13	.14	.15	.15	.17	.18	.16	.18	.20

Empirical distribution found significantly different from the theoretical F distribution at the $p < .05$ level
Empirical distribution found significantly different from the theoretical F distribution at the $p < .01$ level

Table 6: Actual Significance Levels When Using the Nominal Significance of $p < .05$ and Total $n = 60$

Group Sizes Condition A	20 : 20 : 20			18 : 20 : 22			16 : 20 : 24			14 : 20 : 26			12 : 20 : 28		
Group Sizes Condition B	20 : 20 : 20			22 : 20 : 18			24 : 20 : 16			26 : 20 : 14			28 : 20 : 12		
Smallest to Largest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A															
Distributional Shape															
Normal	.05	.05	.05	.04	.04	.05	.04	.03	.04	.04	.03	.03	.03	.03	.03
Platykurtic	.05	.05	.05	.04	.04	.05	.04	.03	.03	.03	.03	.03	.04	.03	.03
Leptokurtic	.04	.05	.05	.04	.04	.05	.04	.03	.03	.03	.03	.03	.03	.03	.03
Moderate Skew	.04	.04	.05	.04	.05	.05	.04	.03	.04	.03	.03	.03	.03	.03	.03
Mod. Skew & Platy	.04	.04	.05	.04	.05	.05	.03	.03	.03	.03	.03	.03	.03	.03	.03
Mod. Skew & Lepto	.04	.05	.05	.04	.04	.05	.04	.03	.04	.03	.03	.03	.03	.03	.03
Ext. Skew & Lepto	.04	.05	.05	.04	.05	.05	.04	.04	.04	.03	.03	.03	.03	.03	.03
ANOVA : CONDITION B															
Distributional Shape															
Normal	.05	.05	.06	.06	.06	.06	.07	.08	.09	.06	.07	.09	.06	.08	.10
Platykurtic	.05	.05	.06	.06	.06	.06	.07	.08	.09	.06	.07	.09	.06	.08	.10
Leptokurtic	.05	.05	.05	.06	.06	.07	.07	.08	.09	.06	.07	.09	.07	.08	.10
Moderate Skew	.05	.05	.06	.06	.06	.06	.06	.08	.09	.06	.08	.09	.06	.08	.10
Mod. Skew & Platy	.05	.05	.06	.06	.06	.06	.07	.08	.09	.06	.07	.09	.06	.08	.10
Mod. Skew & Lepto	.05	.05	.06	.06	.06	.07	.07	.08	.09	.06	.08	.09	.07	.08	.10
Ext. Skew & Lepto	.05	.06	.06	.06	.06	.06	.07	.08	.09	.06	.08	.10	.07	.08	.10
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.05	.05	.06	.04	.04	.04	.03	.03	.03	.04	.04	.03	.04	.04	.04
Platykurtic	.05	.05	.06	.04	.04	.04	.03	.03	.03	.04	.04	.03	.04	.04	.04
Leptokurtic	.05	.05	.06	.04	.04	.04	.03	.03	.03	.04	.03	.03	.04	.04	.04
Moderate Skew	.05	.06	.06	.04	.04	.04	.03	.03	.03	.04	.03	.03	.04	.04	.04
Mod. Skew & Platy	.05	.05	.05	.04	.04	.04	.03	.03	.03	.04	.03	.03	.04	.04	.04
Mod. Skew & Lepto	.05	.05	.06	.04	.04	.04	.03	.03	.03	.04	.03	.03	.04	.04	.04
Ext. Skew & Lepto	.05	.06	.06	.04	.04	.04	.03	.03	.03	.04	.03	.03	.04	.04	.04
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.05	.05	.05	.06	.06	.06	.07	.08	.08	.07	.08	.09	.07	.08	.10
Platykurtic	.05	.05	.05	.06	.08	.06	.07	.08	.08	.07	.09	.09	.07	.08	.10
Leptokurtic	.05	.06	.06	.06	.06	.06	.07	.07	.08	.07	.08	.09	.07	.08	.10
Moderate Skew	.05	.05	.05	.05	.06	.06	.07	.08	.08	.06	.08	.09	.07	.09	.10
Mod. Skew & Platy	.05	.05	.05	.05	.06	.06	.07	.07	.08	.06	.08	.09	.07	.08	.10
Mod. Skew & Lepto	.05	.05	.05	.05	.06	.06	.06	.07	.08	.06	.08	.09	.07	.08	.10
Ext. Skew & Lepto	.06	.06	.06	.05	.05	.06	.07	.07	.08	.06	.08	.09	.07	.08	.11
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.04	.05	.05	.04	.05	.05	.04	.04	.04	.05	.04	.04	.04	.03	.03
Platykurtic	.04	.04	.05	.04	.05	.05	.04	.04	.04	.04	.04	.04	.04	.03	.03
Leptokurtic	.04	.04	.05	.04	.04	.05	.04	.05	.04	.05	.04	.04	.04	.04	.03
Moderate Skew	.04	.04	.05	.05	.05	.05	.04	.04	.04	.04	.04	.04	.04	.03	.03
Mod. Skew & Platy	.04	.04	.04	.05	.05	.05	.04	.04	.04	.04	.04	.04	.04	.04	.03
Mod. Skew & Lepto	.04	.04	.05	.04	.05	.05	.04	.04	.04	.05	.04	.04	.04	.03	.03
Ext. Skew & Lepto	.04	.04	.04	.05	.05	.05	.04	.04	.04	.05	.04	.04	.04	.03	.03
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.05	.05	.05	.06	07	.07	.07	.08	.08	.08	.10	.11	.09	.11	.13
Platykurtic	.05	.05	.05	.06	.07	.07	.07	.08	.08	.08	.10	.11	.09	.11	.13
Leptokurtic	.05	.05	.05	.06	.06	.07	.07	.07	.09	.08	.10	.11	.09	.11	.13
Moderate Skew	.05	.05	.06	.06	.06	.07	.07	.08	.09	.09	.10	.11	.10	.11	.12
Mod. Skew & Platy	.05	.05	.06	.06	.07	.06	.07	.08	.09	.08	.10	.11	.09	.11	.13
Mod. Skew & Lepto	.05	.05	.05	.06	.07	.07	.07	.07	.08	.09	.10	.11	.10	.11	.13
Ext. Skew & Lepto	.05	.06	.06	.07	.07	.07	.07	.08	.08	.09	.11	.12	.10	.11	.13

Empirical distribution found significantly different from the theoretical F distribution at the $p < .05$ level
Empirical distribution found significantly different from the theoretical F distribution at the $p < .01$ level

Table 7: Actual Significance Levels When Using the Nominal Significance of $p \leq .01$ and Total $n = 60$

Group Sizes Condition A	20 : 20 : 20			18 : 20 : 22			16 : 20 : 24			14 : 20 : 26			12 : 20 : 28		
Group Sizes Condition B	20 : 20 : 20			22 : 20 : 18			24 : 20 : 16			26 : 20 : 14			28 : 20 : 12		
Largest to Smallest Group Variance Ratios	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5	1:2	1:3	1:5
ANOVA : CONDITION A															
Distributional Shape															
Normal	.008	.008	.011	.010	.010	.010	.008	.006	.006	.005	.004	.004	.007	.006	.006
Platykurtic	.008	.008	.011	.010	.011	.010	.008	.007	.006	.005	.004	.004	.006	.005	.006
Leptokurtic	.007	.008	.009	.009	.009	.011	.007	.005	.005	.005	.005	.003	.008	.006	.005
Moderate Skew	.008	.009	.010	.008	.010	.011	.005	.006	.007	.005	.004	.004	.007	.006	.005
Mod. Skew & Platy	.007	.008	.008	.007	.010	.013	.007	.006	.007	.005	.004	.003	.006	.005	.005
Mod. Skew & Lepto	.007	.008	.010	.008	.008	.010	.005	.005	.005	.005	.004	.004	.007	.005	.005
Ext. Skew & Lepto	.007	.008	.010	.007	.007	.013	.006	.005	.005	.005	.004	.005	.006	.007	.005
ANOVA : CONDITION B															
Distributional Shape															
Normal	.009	.009	.011	.014	.014	.016	.018	.022	.031	.020	.024	.031	.019	.022	.031
Platykurtic	.009	.009	.011	.013	.014	.016	.019	.023	.033	.019	.023	.030	.020	.023	.031
Leptokurtic	.008	.008	.010	.015	.016	.016	.015	.021	.029	.020	.024	.028	.015	.021	.027
Moderate Skew	.009	.010	.010	.015	.016	.017	.019	.024	.029	.018	.023	.029	.019	.023	.031
Mod. Skew & Platy	.009	.011	.011	.015	.016	.017	.021	.024	.028	.020	.023	.026	.018	.022	.033
Mod. Skew & Lepto	.008	.009	.011	.014	.014	.017	.017	.022	.028	.020	.023	.029	.017	.022	.029
Ext. Skew & Lepto	.008	.009	.010	.014	.016	.018	.017	.023	.028	.019	.022	.027	.019	.025	.031
ANCOVA : CONDITION A, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.017	.016	.016	.008	.009	.011	.007	.007	.006	.009	.006	.005	.011	.008	.007
Platykurtic	.015	.015	.017	.008	.008	.010	.008	.007	.007	.008	.005	.005	.010	.009	.007
Leptokurtic	.015	.015	.016	.007	.007	.010	.006	.006	.006	.006	.006	.005	.010	.008	.007
Moderate Skew	.014	.013	.015	.008	.008	.011	.006	.007	.008	.008	.008	.006	.010	.010	.008
Mod. Skew & Platy	.012	.013	.013	.005	.008	.012	.007	.005	.007	.008	.006	.006	.010	.008	.008
Mod. Skew & Lepto	.015	.016	.016	.007	.006	.008	.007	.005	.005	.007	.007	.006	.010	.008	.007
Ext. Skew & Lepto	.016	.016	.017	.008	.008	.012	.005	.005	.006	.007	.007	.006	.010	.009	.009
ANCOVA : CONDITION B, WITH EQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.010	.010	.010	.011	.013	.015	.015	.020	.026	.014	.021	.029	.017	.025	.034
Platykurtic	.010	.009	.009	.012	.013	.014	.016	.022	.025	.015	.021	.029	.017	.023	.035
Leptokurtic	.009	.010	.011	.015	.013	.014	.015	.017	.023	.015	.020	.027	.016	.020	.029
Moderate Skew	.009	.009	.009	.011	.013	.015	.014	.019	.024	.015	.020	.027	.017	.022	.037
Mod. Skew & Platy	.010	.011	.010	.010	.012	.013	.014	.017	.023	.016	.021	.022	.018	.024	.036
Mod. Skew & Lepto	.009	.010	.011	.012	.012	.012	.013	.017	.021	.013	.018	.027	.015	.022	.031
Ext. Skew & Lepto	.008	.009	.010	.011	.011	.012	.013	.016	.022	.014	.019	.028	.015	.020	.032
ANCOVA : CONDITION A, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.006	.005	.006	.008	.010	.012	.006	.006	.008	.008	.006	.006	.006	.005	.005
Platykurtic	.005	.004	.007	.008	.010	.011	.005	.007	.008	.009	.007	.007	.006	.004	.005
Leptokurtic	.007	.006	.006	.009	.009	.011	.007	.005	.006	.010	.007	.005	.009	.007	.005
Moderate Skew	.006	.004	.004	.008	.010	.011	.003	.004	.006	.009	.008	.008	.007	.005	.006
Mod. Skew & Platy	.004	.004	.004	.008	.009	.014	.003	.004	.006	.009	.009	.008	.006	.006	.006
Mod. Skew & Lepto	.007	.005	.004	.008	.010	.009	.005	.005	.005	.009	.006	.005	.008	.006	.005
Ext. Skew & Lepto	.006	.004	.003	.009	.011	.014	.004	.004	.005	.008	.006	.006	.008	.005	.005
ANCOVA : CONDITION B, WITH UNEQUAL REGRESSION SLOPES															
Distributional Shape															
Normal	.011	.011	.014	.015	.019	.020	.020	.026	.032	.025	.033	.040	.030	.036	.045
Platykurtic	.010	.011	.014	.015	.017	.019	.019	.026	.032	.026	.031	.037	.029	.036	.047
Leptokurtic	.011	.013	.013	.014	.017	.019	.016	.025	.030	.025	.028	.037	.028	.038	.044
Moderate Skew	.011	.012	.014	.016	.017	.020	.021	.028	.032	.025	.031	.040	.031	.036	.043
Mod. Skew & Platy	.011	.013	.015	.016	.019	.019	.023	.026	.031	.024	.029	.037	.029	.035	.044
Mod. Skew & Lepto	.011	.012	.012	.013	.015	.017	.018	.024	.029	.023	.029	.039	.027	.037	.044
Ext. Skew & Lepto	.011	.013	.014	.014	.015	.016	.018	.027	.032	.025	.030	.039	.028	.037	.046

Empirical distribution found significantly different from the theoretical F distribution at the $p < .05$ level
Empirical distribution found significantly different from the theoretical F distribution at the $p < .01$ level

At the nominal significance level of $\alpha < 10$, observed levels of significance for condition A analyses are between .07 and .09, while for condition B they are between .12 and .15. For nominal level $\alpha < .05$, they vary between .03 and .05 for condition A, and between .07 and .09 for condition B. Finally, for the nominal level of $\alpha < .01$, the observed levels vary between .003 and .008 for condition A and between .013 and .033 for condition B (see tables 5, 6 and 7).

Effects of Heterogeneity of Variances in the Design : 14 : 20 : 26

In these analyses, the difference between the smallest and largest group size is 46%. Group differences of this degree result in many significant differences between the empirical distributions and theoretical F. These differences (most at the $\alpha < .01$ level) are found in both condition A and B ANOVA and the ANCOVA with unequal slopes. Interestingly, almost no significant differences emerge with condition A ANCOVA.

At the nominal significance level of $\alpha < 10$, observed levels of significance for condition A analyses are between .06 and .10, while for condition B they range between .12 and .18. For nominal level $\alpha < .05$, they vary between .03 and .05 for condition A, and between .06 and .12 for condition B. Finally, for the nominal level of $\alpha < .01$, observed levels vary between .003 and .010 for condition A and between .013 and .040 for condition B (see tables 5, 6 and 7).

Effects of Heterogeneity of Variances in the Design : 12 : 20 : 28

The difference between the largest and smallest groups in these analyses is 57%. Significant differences between the empirical and theoretical F distributions occur at the $< .01$ level for every ANOVA and condition B ANCOVA that include moderate or extreme heterogeneity of variances. Only a few significant differences emerged among the condition A ANCOVA's ($\alpha <$ .05 level) ; these were usually when extreme heterogeneity of variances was involved.

At the nominal level of $\alpha < 10$, observed levels of significance for condition A analyses are between .05 and .09, while for condition B they range between .12 and .21. For nominal level $\alpha < .05$, they vary between .03 and .04 for condition A, and between .06 and .13 for condition B. Finally, for the nominal level of $\alpha < .01$, observed levels of significance varied between .005 and .011 for condition A and between .015 and .047 for condition B (see tables 5, 6 and 7).

A Final Look at the Simulation Results: A Graphical Analysis

Following significance testing and examination of the actual alpha levels, 274 graphs were drawn which allowed a visual comparison of the differences in observed type I errors that are the result of the different group size combinations. Space does not allow the printing of the graphs in this paper. However, inspection of these graphs revealed some interesting patterns that are masked when analyzing this data purely from an algebraic perspective.

Examining first the ANOVA and ANCOVA graphs with no heterogeneity of slopes, lines representing less than 20% difference between the smallest and largest group size tend to cling together, relatively close to the line representing nominal alpha for condition A analyses. With condition B analyses, those analyses representing group sizes of less than 25% difference tend to cling together as well, although they are not as close to the line representing nominal alpha as the condition A lines for group n's less than 20% different. With both condition A and condition B, when group differences were over 40%, their lines fell far from nominal alpha. Interestingly, the graphs for ANOVA and the graphs for ANCOVA with equal slopes (for almost any specific combination of distributional shape and heterogeneity of variances) appear amazingly similar.

With the ANCOVA graphs when slopes are unequal, however, condition A and B graphs produce patterns that are both unique to their respective conditions and different from one another as well. Lines representing the different group size combinations fell close together and crossed frequently in condition A graphs, suggesting that much of the variation pictured by the graphs is due only to chance variation. These same lines in the condition B graphs fell far from nominal alpha and were spread far from one another as well.

Finally, in comparing the six sets of graphs (ANOVA's for condition A and B, ANCOVA's with equal slopes for conditions A and B, and ANCOVA's with unequal slopes for conditions A and B), it should be noted that the average distance of the lines representing group size are closest to nominal alpha in the condition A ANCOVA's with unequal slopes, while the average distance of those lines is farthest for the condition B ANCOVA's with unequal slopes.

FINDINGS AND CONCLUSIONS

The Complex Nature of the Effects of Violations of Data Set Assumptions

Previous research (e.g., Glass, Peckham and Sanders, 1972; Johnson and Rakow, 1993) suggests that the combination of unequal group variances, unequal group sizes, and (in ANCOVA) unequal group regression slopes pose the greatest threat to robustness. The results of the present study produce findings consistant with previous literature: in those analyses having homogeneity of group variances, the only simulations that emerged as statistically significant from theoretical F were those which had both a large degree of difference in group n's (50% and 57% difference between the largest and smallest groups) and unequal regression slopes. Additionally, among those simulations with near equal n's (the balanced designs, as well as those with group differences of 12.5% and 18%) no significant differences emerged, even when heterogeneity of group variances and regression slopes were present.

Interestingly, unequal sample sizes did not - in and of themselves - threaten the robustness of the ANOVA or ANCOVA results, even when these differences were extreme.

However, unequal n's leave one's research vulnerable to the effects of violations of data set assumptions, a risk that grows greater as the degree of differences in group sizes increase.

The joint relationship among unequal group sizes, heterogeneity of group variances and heterogeneity of regression slopes - crucial to understanding the issues involved here - is a complex one. To fully understand this relationship, one must first understand the pairs of relationships involved.

Looking first at the paired relationship between unequal n's and unequal group variances, when the smallest group is coupled with the smallest variance and the largest group coupled with the largest variance (i.e., a match between magnitude of group sizes and variances), there will result less type I errors than one would expect based on normal theory. However, when the smallest n is coupled with the largest variance and the largest n is coupled with the smallest variance (i.e., an inverse relationship between magnitude of group sizes and variances), the result is more type I errors than expected. In regards to the paired relationship between unequal n's and unequal group regression coefficients, when the smallest n is coupled with the smallest r (i.e., a match between the magnitude of sample sizes and r), more type I errors than expected will emerge. However, when the smallest n is coupled with the largest r (an inverse relationship between magnitude of group sizes and r), there will be less type I errors than expected (e.g., Johnson and Rakow, 1993; McClaren, 1972; Shields, 1978).

Two distintive patterns emerge as these paired relationships are logically extended to a more complex form. The first pattern (condition A in these simulations) occurs when a match between the magnitude of group sizes and variances is coupled with a match in the magnitude between group sizes and r. The former will increase the risk of type I error, while the latter will act to decrease type I errors. The net effect is a wash-out. This phenomenon explains the otherwise baffling pattern that emerged in the condition A ANCOVA simulations involving unequal slopes: even when extremely unequal group sizes were coupled with extreme heterogeneity of variances, statistically significant differences were almost nonexistant. Similarly, the observed significance levels were very close to the alpha levels posited by normal theory, even in face of these data set violations.

The second pattern (condition B) occurs when the inverse coupling of group sizes and magnitude of r is joined with the matched group sizes and variances. Both of these paired combinations will individually increase the risk of type I error. When these two pairs of conditions are combined into a single, compound condition, the risks are additive - and the result is type I error rates inflated far beyond what either of the two pairs of conditions alone could have produced. This, of course, explains why the observed significance levels of the condition B ANCOVA analyses with unequal slopes in this study are so much higher than nominal alpha.

Actual vs. Expected Type I Errors

Typically, when a researcher is engaged in hypothesis testing, the first choice he or she will make is which statistical procedure, from the many available, will be used to do the testing. This decision is quickly followed by a second: which level of significance to use. The level of significance, of course, defines the maximum risk of making type I error that a researcher is willing to accept. When chosing a significance level, the researcher assumes - often on the basis of blind faith alone - that it represents the true risk of making such error.

These results suggest that this faith may not always be well placed. Unequal group sizes alone do not seem to significantly affect type I error rates, however unequal n's coupled with heterogeneity of variances and/or heterogeneity of regression slopes often will. This, in turn, can lead to misleading or inaccurate interpretations of one's data.

Besides simply documenting the results of the simulations in this study, the tables enclosed in this paper serve a second purpose: they offer the research practitioner whose study involves ANOVA or ANCOVA with three groups some idea of the true risk of type I error faced in their own research. This can be particularly useful in situations where the data has already been collected and, upon checking, a researcher finds one or more data set violations. Consulting the table that corresponds with the researcher's chosen nominal level of significance, the sample size combination that best reflects the researcher's own group sizes can be found. From there, the user of the table can locate the particular combination of violations found in his or her data set. The proportions found in the body of the table represent the actual levels of significance (i.e. the observed proportion of type I error made in these simulations) for the nominal significance levels of $\alpha < .10$, $\alpha < .05$ and $\alpha < .01$. While these proportions will not provide the exact risk of type I error for another data set, they should provide estimates more accurate than normal theory provides - at least when three groups with similar group sizes to those in the tables are used.

A word of caution is called for. These tables should be helpful for those researchers using three groups with group size distributions similar to those included on the tables. However, the ability of these proportions to generalize to research not having three groups, or to group size distributions far different from those reflected in the table has not been established. Indeed, research by Tomarken and Serlin (1986) suggests that these proportions may not, in fact, generalize beyond the three group research model. Their research suggests that the effects of heterogeneity of group variances increases as the number of groups increases.

So How Unequal Can Sample Sizes Be, Anyway?

Based on these simulation results, the following general recommendations can be made for the researcher who uses ANOVA or ANCOVA to test for differences between three groups. First, although in theory unequal n's, in and of themselves, do not jeopardize robustness, in practice this will never be the case. The reason is that it is virtually impossible for a researcher to stumble into a research situation where the data contains no violations of statistical assumptions (Glass et al., 1972). Unequal n's will generally exaggerate the effects of any data set violations that do exist.

So how unequal can group n's be? When the difference between the largest and smallest groups is no more than 24%, the results produce type I error rates almost as close to nominal alpha as the balanced (equal n) design. Differences between 25% and 33% seem to be marginally adequate if the smallest group has the smallest variance and the largest group has the largest variance; otherwise group size differences of this magnitude will significantly affect type I error rates. Finally, differences above 33% produce actual type I error rates too different from nominal alpha to be considered statistically robust. Samples this unequal should be avoided if at all possible.

BIBLIOGRAPHY

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I: Effect of inequality of variance in the oneway classification. Annals of Mathematical Statistics, 25, 290 - 302.

Cochran, W.G. (1957). Analysis of covariance: Its nature and uses. Biometrics, 13, 261 - 281.

Elashoff, J.D. (1969). Analysis of covariance: A delicate instrument. American Educational Research Journal, 6, 383 - 401.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods empolyed in the American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Evans, S.H. & Anastasio, E.J. (1968). Misuse of analysis of covariance when treatment effects and covariate are confounded. Psychological Bulletin, 69, 225 - 234.

Feldt, L.S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika, 23, 335 - 353.

Fleishman, A.I. (1978). A method for simulating non-normal distributions. Psychometrika, 43, 521-532.

Games, P.A. & Lucas, P.A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data.

Glass, G.V., Peckham, P.D. & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 42 (3), 237 - 288.

Goodwin, L.D. & Goodwin, W.L. (1988). Statistical techniques in AEJR articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Halpin, G. & Halpin, G. (1988). Evaluation of research and statistical methodologies. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY, November.

Johnson, C. C. & Rakow, E. A. (1993). The effects of violations of data set assumptions when using the oneway, fixed-effects analysis of variance and the one concomitant analysis of covariance. Research in the Schools, in press.

Loftin, L. & Madison, S. (1991). The extreme dangers of covariance corrections. In B.Thompson (Ed.) Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 133-147). Greenwich, CT: JAI Press.

McClaren, V.R. (1972). An investigation of the effects of violating the assumption of homogeneity of regression slopes in the analysis of covariance model upon the F-statistic. Unpublished doctoral dissertation, North State Texas University, Denton, TX.

McLean, J.E. (1974). An empirical examination of analysis of covariance with and without Porter's adjustment for a fallible covariate. Unpublished doctoral dissertation, University of Florida.

McLean, J.E. (1979). The Care and Feeding of ANCOVA. Paper presented at the annual meeting of the Mid-South Educational Research Association in Little Rock, AR, November.

McLean, J.E. (1989). ANCOVA: A Review, Update and Extension. Paper presented at the annual meeting of the Mid-South Educational Research Association in Little Rock, November.

Raaijmakkers, J.G. & Pieters, J.P.M. (1987). Measurement error and ANCOVA: Functional and structural relationships. Psychometrika, 52, 4, 521 - 538.

Scheffe´, H. (1959). The analysis of variance. New York: John Wiley & Sons.

Shields, J.L. (1978). An empirical investigation of the effect of heteroscedascity and heterogeneity of variance on te analysis of covariance and the Johnson-Neyman technique. Army Technical Paper 292, Project no. 2Q762722A777, July.

Student, (W.S. Gossett). (1908). The probable error of the mean. Biometrika, 6, 1 - 25.

Thompson, B. (1992). Misuse of ANCOVA and related "statistical control" procedures. Reading Psychology, 13, iii-xviii.

Tomarken, A., & Serlin, R. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin, 99, 90-99.

Winer, B.J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.